

MPEG-4 TwinVQ による誤り耐性スケーラブル符号化

守谷健弘、森岳志、岩上直樹、神明夫

NTT 研究所

180-8585 武蔵野市緑町 3-9-11

e-メール: moriya.takehiro@lab.ntt.co.jp

ISO/IEC MPEG-4 オーディオの TwinVQ (transform-domain weighted interleave vector quantization : 変換領域重みづけインターリーブベクトル量子化) 楽音符号化と EP (error protection : 誤り保護) ツールを使った誤り耐性スケーラブル符号化の設計を提案する. 16 から 32 kbit/s の符号化実験で, TwinVQ 本来の誤り耐性, 効果的な不均一訂正手法さらに階層のスケーラブル符号化構成の組み合わせにより, 誤り率の高い環境でも劣化を小さくできることを確認した. この符号化枠組みは無線も含む各種のネットワーク, 各種のアプリケーションに柔軟に適用可能である.

An Error-Robust Scalable Coder Based on MPEG-4 TwinVQ

Takehiro Moriya, Takeshi Mori, Naoki Iwakami, and Akio Jin

NTT Laboratories

3-9-11 Midori-cho Musashino, Tokyo, 180-8585 Japan.

e-mail: moriya.takehiro@lab.ntt.co.jp

This paper proposes a design for an error-robust and scalable audio coder based on TwinVQ (transform-domain weighted interleave vector quantization) and EP (error protection) tools defined in ISO/IEC MPEG-4/ Audio. By combining the inherent error-robustness of TwinVQ, the un-equal error-protection scheme, and the hierarchical scalable coding structure, the distortion is confirmed to be very minor, even in high-error channel conditions at bit rates of 16-32 kbit/s. These structures offer a flexible coding framework that can be used for various types of network/storage environments and applications.

1 はじめに

ISO/IEC MPEG-4 オーディオ標準は次世代のマルチメディアコンテンツの符号化として期待され、第1版は1999年に正式に成立し、第2版も仕様が固まった段階にある [1, 2]. インターネットや携帯半導体メモリープレーヤーでの利用を中心に、音楽をMP3 (MPEG 1,2 レイヤー III) の形式で [3] 符号化することが急速に広まっているが、多様な用途には十分対応できない。これに対し、MPEG-4 オーディオは圧縮率向上による基本性能の改善だけでなく、ビットレートやサンプリング周期の拡張、さらにスケーラブル符号化、誤り耐性、再生速度・ピッチの制御などの機能が強化されている。スケーラブル符号化とは、圧縮ビット列の一部からだけでも信号が再生できる機能で、ネットワークや端末の環境に適応したサービスが可能になる。

本報告ではMPEG-4 オーディオ規格のうち、特にTwinVQ 符号化方式のビットレートスケーラブル機能と誤り耐性の特徴に重点をおいた設計を提案する。

なおTwinVQ [4, 5, 6, 7] は1994年にNTTで開発され、平行してMPEG-4の標準化に提案された。ほぼ5年間にわたる標準化の過程で、改良、機能拡張、MPEG-2/4 AAC (Advanced Audio Coder) とのモジュールやインターフェースの共通化が図られた。公式主観品質評価を経て、特に低ビットの楽音符号化方式として正式に採用された。 [8].

2 情報源符号化

2.1 楽音符号化の基本構成

第1図はMPEG-4 オーディオ (第1版) の楽音 (general audio) 符号化の基本枠組みを示す。TwinVQ は AAC と同じ MDCT (Modified Discrete Cosine Transform) による変換符号化の枠組みを使うので、変換に関連する AAC のツールを選択して組み合わせて使うことができる。対象となる AAC のツールは TNS (Time domain Noise Shaping : 時間領域雑音整形), LTP (Long Term Predictor : 長時間予測), 適応 MS (Middle-Side : 中央両側または和差) ステレオ符号化である。ビットレートスケーラビリティは MDCT 係数の量子化を階層的に行うことで実現する。TwinVQ や CELP 方式は

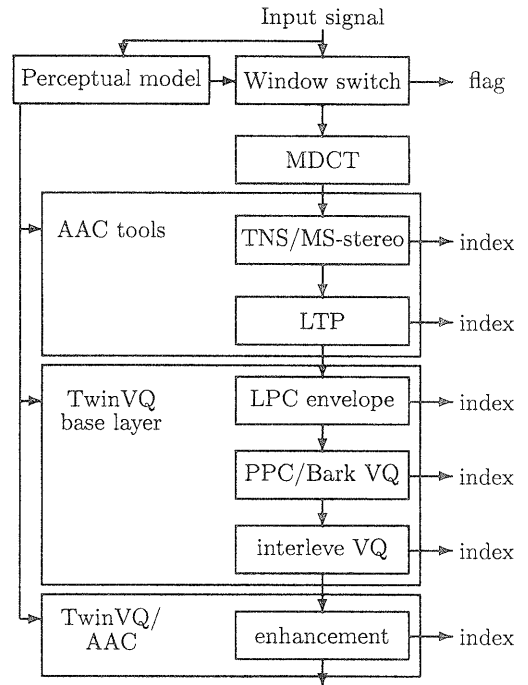


図1: MPEG-4 楽音符号化 (AAC, TwinVQ) の符号器の構成

自分自身で階層量子化が可能であるし、AAC のスケーラブル符号化と組み合わせてそのコア (下位) の量子化として使うこともできる。

2.2 TwinVQ

TwinVQ における MDCT 係数に対する処理は大きく2つに別けられる。ひとつはスペクトル平坦化 (正規化) 処理で、もうひとつは重みづけインターリーブ・ベクトル量子化である。スペクトル平坦化では LPC (Linear Predictive Coding : 線形予測符号化), スペクトル包絡推定, PPC (Periodic Peak Component : 周期的ピーク成分) 抽出, バーク尺度スペクトル推定, パワー正規化の処理を行う。一連の処理の結果、MDCT 係数はほぼ平坦な振幅をもつベクトルとなる。また平坦化に使ったスペクトル情報は聴覚的な変形を加えて、重みつき量子化の重み係数として利用する。LPC 係数の量子化として、音声符号化でほぼ例外なく使われている LSP (Line Spectral Pair : 線スペクトル対) パラメータ [9] を使っている。

重み付けインターリーブベクトル量子化モジュールでは、平坦化した MDCT 係数を下位の周波数からインターリーブしたあとで副ベクトルに分割し、副ベクトルに対して均一な情報を割り当ててベクトル量子化を行う。すなわち、あらかじめ用意しておいた符号帳の中の再生候補ベクトルの中から最もひずみが小さくなるベクトルを選択して、そのインデックスを伝送情報とする。時間領域に再生したときに生じる最終的な量子化ひずみを小さくするためにスペクトル包絡の重みのついたひずみ尺度を用い [10]、さらに 2 つの再生ベクトルの和を出力とする 2 チャンネル共役ベクトル量子化 [12] を用いている。この共役ベクトル量子化は最適化した符号帳ベクトルを用いることで、ひずみを殆ど増加させないで、符号帳のメモリ量を削減できる利点がある。また予備選択を使うことで探索処理を軽減でき、同時に符号誤り耐性を強めることが可能である。

なおベクトル量子化は MDCT 係数だけでなく TwinVQ のなかのパラメータの量子化に幅広く利用されている。LSP パラメータは移動平均型フレーム間予測多段ベクトル量子化 [11] を用いている。PPC には MDCT 係数と同じ構造の重みづけインターリーブ 2 チャンネル共役ベクトル量子化を用いている。バーク尺度包絡には移動平均型フレーム間予測重み付けインターリーブベクトル量子化を用いている。

2.3 ビットレート・スケーラビリティ

ビットレート・スケーラビリティは圧縮ビット列の一部からでも、ビットレートにみあった音が再生できる機能で、ひとつのビット列で、端末の利用者の要求、誤りや混雑などの伝送路の状態、コンテンツや伝送のコストなどに柔軟に対応できる。TwinVQ ではこの機能を MDCT 係数の階層的な多段量子化で実現する [13]。最初の階層（ベース層）では主に低域の MDCT 係数のみを量子化し、その再構成 MDCT 係数と原信号の MDCT 係数の誤差を次の階層（エンハンス層）で量子化する。エンハンス層は 8 階層まで可能で、階層が高くなるにつれてひずみを小さくできるとともに再生信号の帯域を拡大する。

階層符号化では各階層ごとに別のビット列を持つことになるので、極端な場合には階層ごとに物理的に異なる媒体、異なるプロトコルを使うことも可

能となる。このため MPEG-4 の標準に準拠しつつ、きわめて多様な用途に柔軟に利用できる。表 1 にスケーラブル符号化の簡単な使い分け例を示す。

表 1: スケーラブル符号化の利用例

	ベース層	エンハンス層
ビットレート	低い	高い
チャンネル数	実質モノラル	ステレオ
帯域幅	狭い	広い
伝送形式	1 対多 ストリーミング	1 対 1 ダウンロード
料金	無料	有料
符号化方式	TwinVQ	TwinVQ/AAC

3 誤り耐性

3.1 不均一誤り訂正

伝送ビット誤りを訂正する手法として、情報ビットの重要性や波形に及ぼす影響に適應させて強度の異なる訂正を行う、不均一誤り訂正手法があり、デジタル携帯電話などに広く使われている [14, 15, 16]。TwinVQ の圧縮ビット列はパラメータなどの補助情報と MDCT 係数のインデックスの主情報に分離できるので、パラメータ毎に誤りの影響を調べて不均一訂正手法を利用することができる。補助情報には AAC のツールを利用するか否かのフラッグ、窓の形状、スペクトル包絡情報、パワー情報が含まれる。また TwinVQ は従来の多くの楽音符号化と異なり、符号誤りに強い構造になっていることも重要である。すなわち、符号誤りの影響が大きいパラメータ毎またフレームにまたがる適応ビット割り当て、可変長符号を使っておらず、誤りの影響が伝播する再帰型フレーム間予測を使っていない。MDCT 係数の符号帳ベクトルは誤りの影響が少なくなるような設計やインデックス割り当てが行われている。

第 2 図から第 5 図は TwinVQ の圧縮ビット列の各ビット毎の誤り感度を示している。具体的には誤りのないビット列から再生した波形を基準としたときの、誤りを強制的に挿入した波形のひずみを測定している。横軸はビットの番号で、特定のビットに対し、2 フレームに 1 度だけ強制的にビットを反転させている。

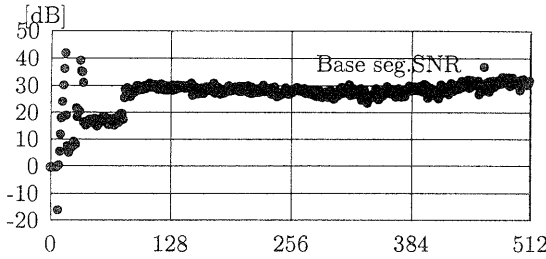


図 2: 誤り感度の例 (ベース層, SNR)

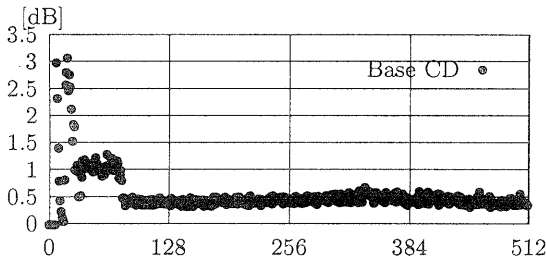


図 3: 誤り感度の例 (ベース層, CD)

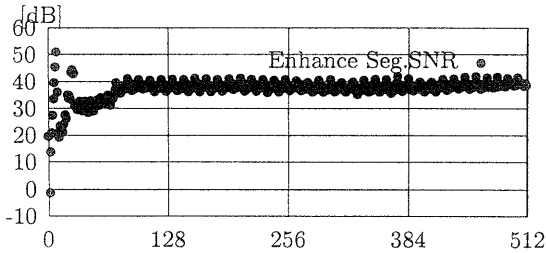


図 4: 誤り感度の例 (エンハンス層, SNR)

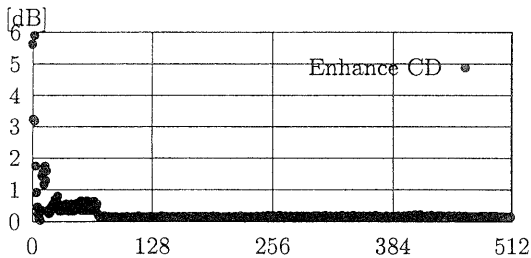


図 5: 誤り感度の例 (エンハンス層, CD)

フレーム毎に SNR を算出して平均したセグメンタル SNR と CD (ケプストラム距離) を用いた。CD は LPC ケプストラム Cep に基づくスペクトル包絡のひずみを測定するもので以下のように定義される。

$$CD = 10 \ln(10) \sqrt{2 \sum_{i=1}^{np} (Cep_A - Cep_B)^2} \quad (1)$$

ここで np は LPC ケプストラムの次数でたとえば 32 である。なお 14496-4 (MPEG-4 コンフォーマンス試験法) でも標準復号器の基準出力と試験復号器の出力波形をセグメンタル SNR と CD で評価することが規定されている。具体的には基準出力と試験出力のセグメンタル SNR が 30dB 以上で、CD が 1dB 以下であれば、二つの波形は実用的に同等とみなせるということである。

この基準と図 2 から 5 までを比べると、全体の 15% ほどを占める左から 70 ビットまでは補助情報で、誤りの影響が SNR や CD に大きく現れているが、大部分を占める主情報に対しては影響が非常に小さいことがわかる。主情報だけをとりあげればフレームあたり 512 ビットで 2 フレームに 1 回の誤りすなわち 1024 ビットに 1 回程度の (10^{-3}) 誤りではセグメンタル SNR が 30dB 以上で、CD が 1dB 以下あり、特に訂正しなくても殆ど品質は劣化しないことが予想できる。また、エンハンス層はベース層の誤差を量子化することもあり、誤りに対する影響が非常に小さい。またエンハンス層の本来の機能として、エンハンス層が完全になってもベース層だけでも、ビットレートに見合う音が再生できる。

3.2 EP(誤り保護) ツール

MPEG-4 オーディオ第 2 版には各種の情報源符号化と柔軟に組み合わせることのできる誤り検出・訂正符号化の枠組みが定義されている。CRC (Cyclic Redundancy Check), SRPCP (Systematic Rate-Compatible Punctured Convolutional) 符号、リード・ソロモン符号が用意され、非常に柔軟に冗長符号の強さを選択できる。またバースト誤りによる誤りの集中を緩和するためのインターリーブ機能なども含まれている。

本報告ではこのツールを表 2 と表 3 に示すビット割り当てで利用する。MPEG-4 オーディオ第 2

表 2: 符号化条件

標本化周波数	32	[kHz]
チャンネル	モノラル	
ベース層 (情報源)	16	[kbit/s]
(冗長)	2.75	[kbit/s]
エンハンス層 (情報源)	16	[kbit/s]
(冗長)	2.75	[kbit/s]
フレーム長	1024	
フレームあたりビット数	512	[bit]

表 3: 訂正符号の割り当て例

階層	クラス	情報源 [bits]	検出 [bits]	訂正 [bits]
ベース	0	12	4	16
	1	22	9	43
	2	476	0	0
エンハンス	0	8	3	22
	1	41	5	58
	2	463	0	0

版ではフレーム内のほぼ重要ビット順にビットが並び替えられており、先頭からクラス 0, クラス 1, クラス 2 に分類している。エンハンス層はそのあとのクラスに分類している。

3.3 コンシールメントと品質

激しいバースト誤りの場合、局所的にフレーム内の多くのビットが壊れるので強力な訂正符号でも訂正できない場合が多く発生する。このような場合、CRC で訂正失敗を検出し、聴感的なひずみを小さくするようなコンシールメントを行う。クラス 0 の CRC が誤りを検出した場合にはフレーム全体のビット構成やパワーの情報が不明となるのでそのフレームの復号波形は激しいひずみを生じる可能性がある。このためこのフレームの復号をあきらめ、前のフレームの波形の振幅を小さくした波形を外挿する。クラス 1 の CRC のみが誤りを検出した場合は対応するパラメータのみを前のフレームからコピーして波形を再生することができる。ただし

この場合でも、MDCT 係数のインデックスにも誤りが含まれる可能性が高いので振幅は小さくする。

簡易な主観評価では今回設計した 16 kbit/s と 32 kbit/s のスケーラブル符号化は符号誤りのない場合はスケーラブルでない同じビットレートの MP3 と同等以上であった。またベース層は 10^{-3} 、エンハンス層は 5×10^{-3} の激しいバースト誤りを付加しても劣化はわずかであった。ベースとエンハンスに独立の伝送路を想定すると、同時に誤りが生ずる可能性がきわめて低くなるので、スケーラブル符号化の品質は 1 層だけの通常の符号化より有利である。なお MP3 のビット列には誤りによる影響が大きいビットが多く、また効果的な不均一誤り訂正付加することは非常に難しい。

4 冗長構成スケーラブル符号化

誤り訂正やコンシールメントを行っても激しいバースト誤りによるひずみは避けられない。このような場合に、スケーラブル符号化の構成を利用して工夫が可能である。通常のスケーラブル符号化ではベース層は原信号 $x(n)$ を量子化し、次のエンハンス層ではベース層での局部復号信号 $Q_0(x(n))$ との量子化誤差 $y(n)(= x(n) - Q_0(x(n)))$ を量子化する。この場合ベース層のビットが復号器で得られないと大きな損失となる。ところがベース層で 1 以下の定数 α 倍した $\alpha x(n)$ を量子化し、エンハンス層で $(x(n) - Q_0(\alpha x(n)))$ を量子化するとエンハンス層にも情報が分散し、ベース層がないときの影響は軽減される。さらに多くのエンハンス層に分散させ、検出符号をと組み合わせると別の伝送路を使うと誤りのない層だけを加算して信号を作ればバースト誤りの影響を大幅に軽減できる。これは簡単な一例に過ぎないが必要に応じた冗長構成が設計できる。なお上記のような組み合わせ例でもビット列と復号器の観点から依然として MPEG-4 オーディオ準拠であることに変わりはない。

5 結び

ISO/IEC MPEG-4 オーディオ標準に準拠した TwinVQ 符号化と EP ツールの不均一誤り訂正を使った、誤り耐性のあるスケーラブル楽音符号化の設計を提案した。16 から 32kbit/s の符号化実験で、

TwinVQ 本来の誤り耐性やスケーラブル符号化の階層構造、さらに不均一訂正を使って 20%程度の冗長ビットを付加すれば、現実に近い誤り条件でも品質の劣化を小さく抑えることを確認した。また誤りのない条件では、スケーラブルでない MP3 相当以上の品質であった。これは設計の一例であり、MPEG-4 オーディオに準拠した枠組みでスケーラブル符号化と誤り制御を組み合わせることで、さらに柔軟で実用性の高い符号化が実現できる。

謝辞 日ごろ、ご指導ご議論いただく金子孝夫リーダ、山森和彦プロジェクトマネージャーならびに研究グループのメンバー、池田和永担当課長をはじめとする NTT 東日本の関係各位に感謝する。

参考文献

- [1] ISO/IEC 14496-3 Information Technology: Coding of Audio-Visual Objects (MPEG-4 Audio).
- [2] ISO/IEC 14496-3/AMD1 Information Technology: Coding of Audio-Visual Objects (MPEG-4 Audio version 2).
- [3] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," *J. Audio Eng. Soc.*, vol. 42, No. 10, pp. 780-792, 1994.
- [4] 岩上, 守谷, 三樹, 池田, 神, "周波数重み付けインターリーブベクトル量子化 (TwinVQ) による楽音符号化," *電子情報通信学会論文誌 A*, Vol. J80-A, No. 5, pp. 830 - 837, 1997.
- [5] N. Iwakami, T. Moriya and S. Miki: "High-quality Audio Coding at less than 64 kbit/s by Using TwinVQ," *Proc. ICASSP'95*, pp. 937-940, 1995.
- [6] T. Moriya, N. Iwakami, K. Ikeda and S. Miki: "Extension and Complexity Reduction of TwinVQ Audio Coder," *Proc. ICASSP'96*, pp. 1029-1032, 1996.
- [7] T. Moriya, N. Iwakami, A. Jin, K. Ikeda and S. Miki: "A Design of Transform Coder for Both Apeech and Audio Signals at 1 bit/sample," *Proc. ICASSP'97*, pp. 1371 - 1374, 1997.
- [8] J. Herr, E. Allamanche, K. Brandenburg, M. Dietz, B. Teichman, G. Grill, A. Jin, T. Moriya, N. Iwakami, T. Norimatsu, M. Tsushima, and T. Ishikawa: "The integrated Filterbank Based Scalable MPEG-4 Audio Coder," *105th AES Convention Preprint*, NO. 4810, 1998.
- [9] F. Itakura, T. Kobayashi and M. Honda, "A Hardware Implementation of a New Narrow to Medium Band Speech Coding," *Proc. ICASSP'82*, pp. 1964-1967, 1982.
- [10] T. Moriya and M. Honda, "Transform Coding of Speech Using a Weighted Vector Quantizer," *IEEE JSAC*, vol. 6, pp. 425-431, 1988.
- [11] A. Kataoka, T. Moriya and S. Hayashi, "An 8 kbit/s Speech Coder Based on Conjugate Structure CELP," *Proc. ICASSP'93*, pp. II-592-595. 1993.
- [12] T. Moriya, "Two-channel Conjugate Vector Quantizer for Noisy Channel Speech Coding," *IEEE JSAC*, vol. 10 pp. 866-874. 1992.
- [13] A. Jin, T. Moriya, T. Norimatsu, M. Tsushima, and T. Ishikawa, "Scalable Audio Coder Based on Quantizer Units of MDCT Coefficients," *Proc. ICASSP'99*, pp. II-897 - 900, 1999.
- [14] 池田, 守谷, 岩上, "誤り保護付き TwinVQ 楽音符号化方式," *電子情報通信学会論文誌 D-II*, Vol. J80-D-II, No. 5, pp. 1016 - 1025, 1997.
- [15] K. Ikeda, T. Moriya N. Iwakami and S. Miki: "Error Protected TwinVQ Audio Coding at less than 64 kbit/s," *Proc. IEEE Speech Coding Workshop*, pp. 33-34, 1995.
- [16] K. Ikeda, T. Mori, T. Moriya, N. Iwakami and T. Kaneko, "Audio Transfer System on PHS Using Error Protected TwinVQ," *IEEE Trans. Consumer Electronics*, Vol.44, pp.1032-1038, 1998.